DOCUMENT RESUME

ED 241 117 JC 840 810

AUTHOR Svinicki, Marilla; Koch, Bill
TITLE The Great Essay/Multiple Choice Debate: Different Strokes for Different Folks.
PUB DATE 3 Feb 84
NOTE 4p.; From the "Center for Teaching Effectiveness Newsletter," v5 n2 Nov 1983.
PUB TYPE Collected Works - Serials (022) -- Guides - Classroom Use - Guides (For Teachers) (052)
JOURNAL CIT Innovation Abstracts; v6 n3 Feb 3 1984

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Comparative Analysis; *Essay Tests; *Multiple Choice Tests; *Test Format; Testing; Test Selection; *Test Theory
IDENTIFIERS PF Project

ABSTRACT

The decision of whether to use essay tests or multiple choice tests depends on several qualifiers related to the different characteristics of the tests and the needs of the situation. The most important qualifier involves matching the type of test to the instructional objectives being tested, with multiple choice tests being used to measure a grasp of basic facts and essay tests being used to measure the ability to communicate facts concisely. Theoretical differences between the tests include: level of cognitive processing and memory required; testing time spent on thinking versus production; sensitivity to specific learning; level of cognitive complexity which is measured; confounded results caused by student ability in communication skills; and student preparation. Technical differences include range of course content which can be tested; test reliability; and capability of analyzing test items for improvement purposes. Two practical considerations are the time involved in producing, administering, and grading the test and the potential for cheating and guessing. It is important to consider using a variety of evaluation vehicles so that students who are stronger in one response type than another will have equal opportunities to perform well. (DC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED241117

The Great Essay/Multiple Choice Debate: Different Strokes for Different Folks

Innovation Abstracts
v6 n3 February 3, 1984

JC840810

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Karen Watkins

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

✓ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

INNOVATION ABSTRACTS

VOL. VI
NO. 3

Published by the National Institute for Staff and Organizational Development
With support from the W. K. Kellogg Foundation

THE GREAT ESSAY/MULTIPLE CHOICE DEBATE: DIFFERENT STROKES FOR DIFFERENT FOLKS

There has been a long-standing debate on academic testing--which is better, essay testing or multiple choice testing? Anyone in education as long as we knows that whenever a question is phrased that way, the answer is "neither," followed by a long string of qualifiers. So it should come as no surprise that this is an appropriate answer to our question as well and that our purpose here is to present the qualifiers.

The first and most important qualifier in determining which question type is preferable is the instructional objective being tested. To coin a phrase, "you can't weigh a duck with a yardstick." The measure you choose must match the purpose of the measurement, if not physically, at least conceptually. So if your purpose is to measure the student's grasp of basic facts, the type of test items you choose should require just that and nothing more. If, on the other hand, you want to measure a student's ability to communicate facts concisely, you need to choose a totally different task for the test. None of this is all that startling, but it does require a commitment on the part of the instructor to define exactly what the course objectives are. Then, for the tests to have content validity, there must be a close correspondence between the course objectives, the instructional activities, and the test items used to evaluate the learning that took place. And those course objectives must be stated not just in vague terms like "an understanding of the principle of one man, one vote" but more specifically like "recognizing examples of situations where the one man, one vote principle should be applied" or "predicting how the principle of one man, one vote might affect election outcomes under a variety of conditions" or "proposing solutions to problems of election reform which reflect the one man, one vote principle." These three alternative interpretations of "an understanding of the one man, one vote principle" are all valid; yet they require different types of testing. So the first task of any test designer *must* be to match the instructional objectives to the types of items on the test, laying out a blueprint which charts the course content against the level of understanding desired.

Theoretical Differences

The second set of qualifiers for choosing item types falls under the heading of differences in theoretical properties between the essay item and the multiple choice item. The first of these qualifiers involves the level of cognitive processing and memory required by the two item types. In essay items, the student must use the cues present in the question to search back through his memory and *recall* all the relevant information, use that information to structure a correct response, and write it down. In multiple choice items, the student needs to *recognize* a correct choice either by matching the alternatives to a correct answer that already exists in memory, generating a correct answer and comparing the alternatives to it, or evaluating each alternative as a potential answer and trying to reason from the question to that answer. The advantage of the multiple choice over the essay, from the student's perspective, is that there is an immediate check on memory. If none of the alternatives match his preferred response, he knows that his choice is faulty (except where "none of the above" is an option).

Along similar lines is another, more empirically based characteristic to consider: the testing time spent on *thinking versus production*. Even if complex multiple choice items do indeed require the same initial thought process as essay items, there is a distinct difference in the amount of time for answer production. We often forget that there is simply a certain amount of time needed to physically write the annswer out for an essay, over and above the time needed to compose it. In multiple choice items, the time can be devoted almost exclusively to thinking about the answer and the alternatives. Therefore, an instructor should consider where the student's time would be spent most effectively.

A third characteristic of these two types is their differential sensitivity to learning. Because the amount of information present in the item is greater in a multiple choice item than in an essay item, there is more information which can trigger the memory of the correct response. Therefore, multiple choice items are more sensitive to smaller amounts of specific learning (as evidenced by their susceptibility to cramming). On the other hand, essay items allow the student to throw in everything he knows about a subject; and while these items may be more sensitive to breadth of learning, they are more vulnerable to padding or skirting the question.

A fourth typical characteristic, though not one which is inherent in the item format itself, is the difference in cognitive complexity which is usually demanded by each type of item. In general, essay items tend to be

used to measure more complex learning, such as analysis or synthesis while multiple choice items are used more for basic comprehension and application. It is possible to use multiple choice items to probe for more complex understanding, but it is difficult and time-consuming to write this type of test item.

A fifth characteristic which distinguishes these two types of items is that essay items require more than content knowledge. It has been frequently demonstrated that even the most sophisticated scorer is not immune to the effects of communication skills in evaluating essay answers. The same can be a problem from the student's point of view as well. The confounding of communication skills with content knowledge can be a bane or a boon, depending upon the student's facility with each. The instructor must remember that the confounding exists both at the time of testing and at the time of scoring.

A last theoretical difference between these two types of items may be a psychological one. There is research evidence that there is a difference in the way students view the content and the appropriate way to prepare for an exam depending upon the type of exam being given. Essay exams encourage specificity and concentration on facts. This may be an artifact of the tendency for instructors to use essay items for more general concepts while using multiple choice items for specifics. However, until that trend is reversed, students will continue to approach the two types of exams that way.

Technical Differences

From a measurement point of view, these two types of items are quite distinct. The first difference between the two types lies in their respective abilities to allow for adequate sampling of the content to be tested. Since multiple choice items require less time to answer than do essay items, more items can be included in an exam; therefore, more areas can be tested. The wider the sample range, the more valid and reliable a single test can be.

A second consideration is test reliability. Multiple choice tests, in general, are more reliable than essay tests. This greater reliability is primarily a function of the required grading procedures. Essay items are subject to variability due to errors or influences in grading as well as achievement level differences among the test takers, while the variability of multiple choice items comes mostly from differences among test takers. It is possible to train essay graders to be more consistent and reliable, but such training occurs infrequently.

Finally, multiple choice items lend themselves to statistical analysis for evaluation and improvement purposes. Measurement and evaluation centers can assess the effectiveness of test items or provide instructors with information for completing their own. Although comparable analyses can be done on essay items, the procedures must be done by hand and are far less reliable.

Practical Differences

There are two other practical differences between these two item types. Perhaps the most significant consideration in choosing a test type is that of the time commitment involved in producing, administering and grading the exam. Multiple choice tests are very time-consuming to construct, but they take little time to grade. Also the time during testing is devoted almost exclusively to thought processes rather than production of an answer. Essay tests are less complicated to construct, but they are very time-consuming to grade. During testing, a significant proportion of student time is devoted to production of the answer.

A second consideration is the potential for cheating and guessing. Multiple choice items require that the instructor be particularly alert to both behaviors; they are easily accomplished and difficult to spot. Essay items are much less likely to foster either type of behavior although students may try to "pad" their way through an essay item without knowing the subject.

Making the Choice

It is important to remember that item types should match test objectives. Using a variety of item types serves students well; those who are stronger in one response type than another will have equal opportunities to perform well. Finally, it is important to consider using a variety of evaluation vehicles for assessing student learning. Putting all your instructional eggs in one basket is like trying to document teaching effectiveness on the basis of student evaluations alone.

(From the Center for Teaching Effectiveness *Newsletter*, Vol. 5, No. 2, November 1983)

Marilla Svinicki
Center for Teaching Effectiveness
The University of Texas at Austin

Bill Koch
Measurement and Evaluation Center
The University of Texas at Austin

For further information, contact Marilla Svinicki at the Center for Teaching Effectiveness, 2202 Main Building, The University of Texas at Austin, Texas, 78712.

Suanne D. Roueche, Editor
February 3, 1984, VOL. VI, NO. 3

INNOVATION ABSTRACTS is a publication of the National Institute for Staff and Organizational Development, EDB 348, The University of Texas at Austin, Austin, Texas 78712, (512) 471-7545. Subscriptions are available to nonconsortium members for $35 per year. Funding in part by the W. K. Kellogg Foundation and the Fund for the Improvement of Post-Secondary Education. Issued weekly when classes are in session during fall and spring terms and bimonthly during summer months.

© The University of Texas at Austin, 1984
Further duplication is permitted only by MEMBER institutions for their own personnel.